
Penerapan Algoritma *K-Means* dan *K-Medoids* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Perumahan Rumah Tangga Tahun 2020

Fahriza Rianda^{1*}

¹Prodi D-IV Statistika Politeknik Statistika STIS, Jakarta Timur, Indonesia

* Corresponding author, email: izarian321@gmail.com

Abstract

Statistics Indonesia explained that the percentage of households in Indonesia that had access to decent, safe and affordable housing during the 2019-2020 period was still below 60 percent. Of course, this is a big job for the government to be able to achieve the target in the RPJMN 2020-2024, which is up to 70 percent in providing decent, safe, and affordable housing for the community by 2024. This study aims to group provinces in Indonesia based on indicators. household housing by applying and choosing the best algorithm among *k-means* and *k-medoids*. Based on the selection of the best algorithm, *k-means* is the best algorithm in classifying provinces in Indonesia compared to *k-medoids* with three clusters. The results of the grouping of provinces in Indonesia are expected to assist the government in dealing with problems related to household housing indicators so that the government's target of increasing the percentage of households occupying decent, safe, and affordable housing can be achieved.

Keywords: Household Housing Indicators, Best Algorithm Selection, *k-means*, *k-medoids*.

Abstrak

Badan Pusat Statistik menjelaskan bahwa persentase rumah tangga di Indonesia yang memiliki akses terhadap hunian layak, aman, dan terjangkau selama periode 2019-2020 masih berada di bawah 60 persen. Tentunya hal ini menjadi suatu pekerjaan yang besar bagi pemerintah untuk dapat mencapai target pada RPJMN 2020-2024 yaitu hingga 70 persen dalam menyelenggarakan hunian layak, aman, dan terjangkau bagi masyarakat pada tahun 2024. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator perumahan rumah tangga dengan menerapkan dan memilih algoritma terbaik diantara *k-means* dan *k-medoids*. Berdasarkan pemilihan algoritma terbaik, *k-means* merupakan algoritma terbaik dalam mengelompokkan provinsi-provinsi di Indonesia dibandingkan *k-medoids* dengan jumlah kluster sebanyak tiga. Hasil pengelompokan provinsi-provinsi di Indonesia tersebut diharapkan dapat membantu pemerintah dalam menangani masalah-masalah yang berkaitan dengan indikator perumahan rumah tangga agar target pemerintah untuk dapat meningkatkan persentase rumah tangga yang menempati hunian layak, aman, dan terjangkau dapat tercapai.

Kata Kunci: Indikator Perumahan Rumah Tangga, Pemilihan Algoritma Terbaik, *k-means*, *k-medoids*.

1. Pendahuluan

Kebutuhan dasar manusia untuk dapat hidup layak serta sejahtera tentunya harus didukung oleh kondisi perumahan dan permukiman yang memadai. Salah satu amanat UUD 1945 adalah setiap orang berhak mendapatkan tempat tinggal dan lingkungan hidup yang baik dan sehat. Adanya amanat dari UUD 1945 tersebut menyebabkan negara bertanggung jawab dalam memenuhi kebutuhan bertempat tinggal masyarakat serta

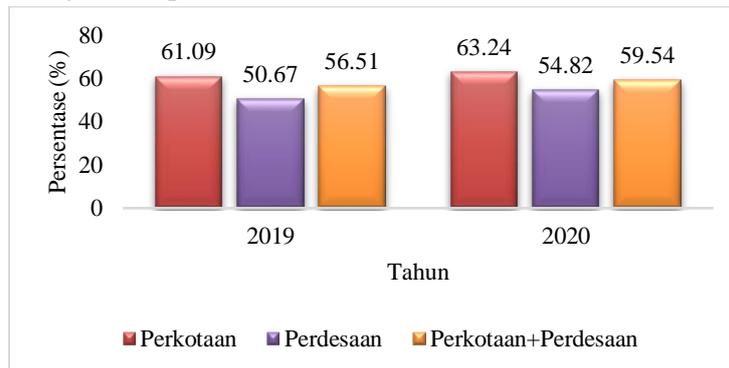
melindungi dan meningkatkan kualitas permukiman dan lingkungan di wilayah masyarakat tersebut agar masyarakat mampu bertempat tinggal di kawasan yang layak, aman, dan terjangkau.

Pemerintah saat ini tengah berupaya untuk meningkatkan akses masyarakat terhadap perumahan dan pemukiman yang layak, aman, dan terjangkau. Sejak tahun 2015, pemerintah melalui Kementerian Pekerjaan Umum dan Perumahan Rakyat telah membuat Program Sejuta Rumah agar masyarakat Indonesia khususnya bagi masyarakat berpenghasilan rendah dapat memiliki rumah yang layak huni [1]. Kemudian pemerintah melalui Rencana Pembangunan Jangka Menengah Nasional (RPJMN) tahun 2020-2024 juga telah menargetkan pada tahun 2024 persentase rumah tangga yang menempati hunian layak, aman, dan terjangkau sebesar 70 persen [2].

Rencana pembangunan perumahan dan permukiman pada RPJMN juga sejalan dengan semangat global dalam *Sustainable Development Goals* (SDGs). Dari 17 aspek tujuan pembangunan berkelanjutan pada SDGs, aspek perumahan merupakan tujuan *ke-11*, khususnya pada indikator 11.1 yaitu menjadikan kehidupan dan kota permukiman berkelanjutan dengan memastikan akses terhadap hunian yang layak, aman, dan terjangkau. Tidak hanya itu, aspek perumahan pada SDGs juga terdapat pada tujuan-tujuan lainnya yaitu pada tujuan *ke-1* mengenai pemenuhan kebutuhan dasar bagi masyarakat miskin serta tujuan *ke-6* mengenai air bersih dan sanitasi yang layak [3]. Dengan demikian, aspek perumahan merupakan aspek yang sangat penting karena menyangkut berbagai aspek kehidupan khususnya kesehatan dan kesejahteraan.

Pandemi Covid-19 yang telah memasuki Indonesia sejak awal tahun 2020 tentunya memberikan banyak dampak, khususnya pada cara menjalani kehidupan. Pandemi Covid-19 yang menyebar begitu cepat di wilayah-wilayah Indonesia membuat pemerintah membatasi aktivitas masyarakat. Sehingga masyarakat melakukan berbagai aktivitas kehidupan di rumah agar terhindar dari pandemi Covid-19. Tidak hanya itu, masyarakat juga harus mempraktikkan perilaku hidup sehat untuk dirinya dan lingkungannya. Adanya aktivitas di rumah saat pandemi Covid-19 ini membuat kebutuhan akan hunian dan lingkungan yang sehat sebagai tempat hidup dan beraktivitas sangat diperlukan. Pemerintah terus berupaya menyediakan tempat tinggal yang layak huni bagi masyarakat untuk bisa meminimalkan masyarakat terhindar dari pandemi Covid-19. Program Sejuta Rumah dari pemerintah telah berhasil membangun 965.217 unit rumah di seluruh Indonesia selama tahun 2020 [4]. Upaya pemerintah ini dirasa cukup berdampak pada peningkatan persentase rumah tangga di Indonesia yang memiliki akses terhadap hunian yang layak dan terjangkau pada tahun 2020 sebesar 3,03 persen dari tahun 2019 yang dikeluarkan oleh Badan Pusat Statistik (BPS) pada gambar 1. Namun BPS (2021) menjelaskan persentase rumah tangga di Indonesia yang memiliki akses terhadap hunian layak, aman, dan terjangkau selama periode 2019-2020 masih berada di bawah 60 persen [5]. Tentunya hal ini menjadi suatu pekerjaan yang besar bagi pemerintah untuk dapat

mencapai target hingga 70 persen dalam menyelenggarakan hunian layak, aman, dan terjangkau bagi masyarakat pada tahun 2024.



Sumber: BPS (diolah)

Gambar 1. Persentase rumah tangga dengan hunian layak, aman dan terjangkau menurut daerah tempat tinggal 2019-2020

Pengelompokan wilayah-wilayah di Indonesia berdasarkan kondisi perumahan rumah tangga sangat diperlukan bagi pemerintah dalam mencapai targetnya untuk menyediakan hunian yang layak, aman, dan terjangkau bagi masyarakat. Pada beberapa penelitian terdahulu, indikator perumahan hanya dibahas secara umum serta digunakan sebagai salah satu atribut dalam mengelompokkan wilayah tertentu berdasarkan indikator seperti kesejahteraan, kemiskinan, ataupun kesehatan. Penelitian dari Wijayanti, Rahmi, dan Yanuar (2021) menggunakan algoritma *fuzzy c-means* untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator kesehatan lingkungan [6]. Pada penelitian tersebut beberapa indikator perumahan seperti air minum layak dan sanitasi layak dimasukkan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator kesehatan lingkungan. Kemudian penelitian dari Latuhimallo, Talakua, dan Leleury (2021) menggunakan algoritma kluster hirarki untuk mengelompokkan kabupaten dan kota di Provinsi Maluku berdasarkan indikator kesejahteraan rakyat [7]. Dalam pengelompokan tersebut, beberapa indikator perumahan yang digunakan adalah jenis atap terluas dan status kepemilikan rumah.

Dengan demikian, berdasarkan penjelasan yang telah disampaikan, penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator perumahan rumah tangga. Indikator-indikator perumahan rumah tangga yang akan digunakan dalam penelitian ini berasal dari BPS. Indikator-indikator yang diperoleh dari BPS tersebut mencakup kriteria rumah layak huni dan status penguasaan tempat tinggal [8]. Namun peneliti membatasi algoritma kluster yang akan digunakan yaitu algoritma *k-means* dan *k-medoids*. Kedua algoritma ini dibandingkan untuk mendapatkan model terbaik dalam mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator perumahan rumah tangga. Hasil penelitian ini diharapkan dapat membantu pemerintah dalam hal menyediakan informasi mengenai provinsi-provinsi yang perlu peningkatan

pada persentase rumah tangga yang menempati hunian layak, aman, dan terjangkau, sehingga target pemerintah dalam RPJMN dapat tercapai pada tahun 2024.

2. Material dan Metode

2.1 *K-Means*

K-means adalah algoritma klusterisasi yang paling banyak digunakan dalam berbagai aplikasi. Hal ini disebabkan algoritma tersebut sederhana untuk dijalankan, relatif cepat, dan mudah beradaptasi [9]. Algoritma *k-means* merupakan algoritma yang berbasis partisi, artinya bahwa algoritma *k-means* akan membagi sekumpulan data ke dalam sejumlah k kluster yang telah ditetapkan di awal. Secara umum, tahapan dari algoritma *k-means* adalah sebagai berikut [10]:

1. Menentukan jumlah kluster k .
2. Menginisiasi k pusat (*centroid*) kluster dengan cara acak artinya pusat-pusat kluster diberi nilai awal dengan angka-angka acak.
3. Mengalokasikan semua data/objek ke kluster terdekat. Kedekatan antara dua objek atau kedekatan objek ke suatu kluster ditentukan berdasarkan ukuran jarak kedua objek tersebut. Pada tahap ini perlu dihitung jarak tiap data ke tiap pusat kluster (*centroid*). Kluster yang terpilih adalah jarak antara titik pusat kluster dengan data/objek yang akan dikelompokkan paling kecil.
4. Menghitung kembali pusat kluster dengan keanggotaan kluster yang ada. Pusat kluster dapat berupa rata-rata dari semua data/objek dalam kluster tertentu.
5. Mengulangi langkah ketiga dan keempat apabila pusat kluster masih dapat berubah. jika pusat kluster tidak berubah, proses pengelompokan (*clustering*) selesai.

2.2 *K-Medoids*

K-medoids atau yang disebut juga dengan algoritma *Partitioning Around Medoids* (PAM) adalah algoritma klusterisasi yang mampu mengatasi kelemahan pada algoritma *k-means* yang sensitif terhadap *noise* dan *outlier* [11]. Algoritma *k-medoids* juga termasuk algoritma yang berbasis partisi. Perbedaan antara algoritma *k-medoids* dan *k-means* terletak pada pemilihan titik pusat yaitu algoritma *k-medoids* menggunakan objek perwakilan sebagai pusat kluster untuk setiap kluster, sedangkan *k-means* menggunakan nilai rata-rata sebagai pusat kluster. Secara umum, cara kerja algoritma *k-medoids* adalah sebagai berikut [12]:

1. Menginisiasi pusat kluster sejumlah k .
2. Mengalokasikan setiap data/objek ke kluster terdekat menggunakan ukuran jarak.
3. Memilih secara acak data/objek pada masing-masing kluster sebagai kandidat *medoid* baru.
4. Menghitung jarak setiap objek yang berada pada masing-masing kluster dengan kandidat *medoid* baru.

5. Menghitung total simpangan dengan menghitung nilai total jarak (*distance*) baru – total jarak lama. Jika total simpangan lebih kecil dari nol, tukar data/objek dengan data klaster baru untuk membentuk sekumpulan k data/objek baru sebagai *medoid*.
6. Mengulangi langkah ketiga hingga kelima hingga tidak terjadi adanya perubahan *medoid*.

2.3 Konsep Jarak Pada *K-Means* dan *K-Medoids*

Perbedaan antara suatu data/objek p di dalam klaster C_i dengan pusat klaster atau disebut juga ketika memakai algoritma *k-means* yaitu *centroid* c_i dihitung menggunakan jarak Euclidean [13]. Jarak Euclidean secara umum dapat didefinisikan sebagai berikut:

$$d(\mathbf{i}, \mathbf{j})^2 = (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2 \quad (1)$$

dimana i dan j adalah dua objek yang memiliki n atribut yang bernilai numerik. Kualitas klaster C_i dengan memakai algoritma *k-means* dapat diukur dengan meminimalkan variasi di dalam klaster, yaitu jumlah kesalahan kuadrat atau *sum of squared error* (E) antara semua objek dalam klaster C_i dan *centroid* c_i yang didefinisikan sebagai berikut:

$$E = \sum_{i=1}^k \sum_{p \in C_i} d(\mathbf{p}, \mathbf{c}_i)^2 \quad (2)$$

Berbeda halnya dengan *k-means*, algoritma tersebut melakukan partisi dengan cara meminimalkan jumlah ketidaksamaan (*dissimilarity*) antara setiap objek p dan objek representatif (*medoids*) o_i yaitu menggunakan jumlah kesalahan mutlak atau *absolute* (E). Ukuran jarak yang diterapkan dalam *k-medoids* biasanya menggunakan jarak Manhattan yang secara umum didefinisikan sebagai berikut:

$$d(\mathbf{i}, \mathbf{j}) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}| \quad (3)$$

dimana $|x_{in} - x_{jn}|$ menyatakan selisih absolut antara nilai atribut ke- n pada objek ke- i dan nilai atribut ke- n pada objek ke- j . Sementara jumlah kesalahan absolut yang disebutkan di atas didefinisikan sebagai berikut:

$$E = \sum_{i=1}^k \sum_{p \in C_i} d(\mathbf{p}, \mathbf{o}_i) \quad (4)$$

Akan tetapi, menurut Kassambara [14], kedua ukuran jarak baik Euclidean maupun Manhattan dapat digunakan pada algoritma *k-medoids*.

2.4 Pemilihan Algoritma Terbaik

Kassambara [14] menjelaskan bahwa untuk membandingkan beberapa algoritma klusterisasi dapat menggunakan ukuran validasi internal yaitu *connectivity*, *dunn index*, dan koefisien *silhouette*.

Connectivity mengukur jarak antara suatu objek yang ditempatkan pada kluster yang sama dengan tetangga terdekat mereka. Nilai *connectivity* terletak pada interval nol sampai tak hingga. Nilai *connectivity* yang minimum memberikan gambaran bahwa kluster memiliki kualitas yang lebih baik. *Connectivity* dapat didefinisikan sebagai berikut [15]:

$$conn = \frac{1}{B} \sum_{i=1}^n \sum_{j=1}^L x_{i,nn_i(j)} \quad (5)$$

dimana L adalah parameter yang memberikan jumlah tetangga terdekat yang digunakan dan B adalah jarak antara masing-masing pusat kluster atau *the between sum of squares*, dan n adalah jumlah observasi. Kemudian $nn_i(j)$ artinya adalah tetangga terdekat *ke-j* dari observasi *ke-i* dan jika $x_{i,nn_i(j)}$ sama dengan nol, *i* dan *j* berada pada kluster yang sama.

Dunn index menghitung validasi kluster berdasarkan ukuran seberapa dekat objek-objek yang ada di dalam kluster atau *cluster cohesion* dan seberapa baik kluster tersebut terpisah dari kluster lainnya (*separation*). Nilai *dunn index* yang besar menandakan bahwa kluster-kluster memiliki kohesi dan terpisah dengan baik. *Dunn indeks* dapat didefinisikan sebagai berikut [15]:

$$Dunn = \frac{\min_{1 \leq i < j \leq q} d(C_i, C_j)}{\max_{1 \leq k \leq q} d(C_k)} \quad (6)$$

dimana $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$ adalah ukuran ketidaksamaan diantara dua kluster dan $diam(C) = \max_{x, y \in C} d(x, y)$ adalah jarak maksimum antara setiap objek di dalam kluster.

Koefisien *silhouette* mengukur seberapa baik suatu objek ditempatkan pada kluster tertentu. Dalam menghitung nilai koefisien *silhouette* dari sebuah objek *ke-i* terdapat dua komponen yaitu a_i yang merupakan rata-rata jarak objek *ke-i* terhadap semua data lainnya dalam satu kluster dan b_i didapatkan dengan menghitung rata-rata jarak objek *ke-i* terhadap semua data dari kluster lain yang tidak dalam satu kluster dengan objek *ke-i* [9]. Semakin besar nilai koefisien *silhouette ke-i* (S_i) yaitu hampir mendekati satu, objek tersebut semakin tepat di dalam sebuah kluster. Sedangkan nilai koefisien *silhouette* yang bertanda negatif menandakan bahwa objek tidak tepat berada di dalam klasternya sebab objek tersebut lebih dekat dengan kluster lain. Koefisien *silhouette* didefinisikan sebagai:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (7)$$

Kassambara [14] menjelaskan untuk menentukan jumlah kluster yang optimal dan mengukur validasi suatu kluster dengan menggunakan algoritma tertentu menggunakan rata-rata dari koefisien *silhouette* pada objek untuk setiap kluster k yang berbeda. Sehingga jumlah kluster k yang optimal dan seberapa baik kluster yang dihasilkan diperoleh dari nilai rata-rata *silhouette* yang maksimal.

2.5 Cakupan Penelitian

Ruang lingkup pada penelitian ini adalah untuk mengelompokkan provinsi di Indonesia berdasarkan indikator perumahan rumah tangga dengan membandingkan algoritma *k-means* dan *k-medoids*. Penelitian ini menggunakan data yang berasal dari BPS. BPS menyediakan indikator-indikator perumahan yang berasal dari Survei Sosial Ekonomi Nasional (SUSENAS) diantaranya mencakup kriteria rumah layak huni dan status penguasaan tempat tinggal [8]. Variabel atau atribut yang digunakan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator perumahan rumah tangga adalah:

Tabel 1. Atribut yang digunakan untuk mengelompokkan provinsi di Indonesia.

No.	Atribut Tahun 2020
1	Persentase rumah tangga menurut provinsi dan status kepemilikan rumah milik sendiri (X1)
2	Persentase rumah tangga menurut provinsi dan luas hunian per kapita $\leq 7,2 \text{ m}^2$ (X2)
3	Persentase rumah tangga menurut provinsi dan memiliki atap terluas bukan ijuk/lainnya (X3)
4	Persentase rumah tangga menurut provinsi dan memiliki dinding terluas bukan bambu/lainnya (X4)
5	Persentase rumah tangga menurut provinsi dan lantai terluas bukan tanah (X5)
6	Persentase rumah tangga menurut provinsi dan sumber air minum layak (X6)
7	Persentase rumah tangga menurut provinsi dan memiliki akses terhadap sanitasi layak (X7)

Sumber: BPS

2.6 Metode Analisis

Penelitian ini menggunakan algoritma kluster berbasis partisi yaitu metode yang bekerja dengan memilih jumlah kluster awal terlebih dahulu kemudian setiap data/objek digabungkan ke dalam kluster-kluster yang telah ditentukan tersebut. Analisis kluster berbasis partisi yang digunakan dalam penelitian ini mencakup algoritma *k-means* dan *k-medoids*. Adapun tahapan pengelompokan menggunakan algoritma *k-means* dan *k-medoids* pada penelitian ini dijelaskan sebagai berikut:

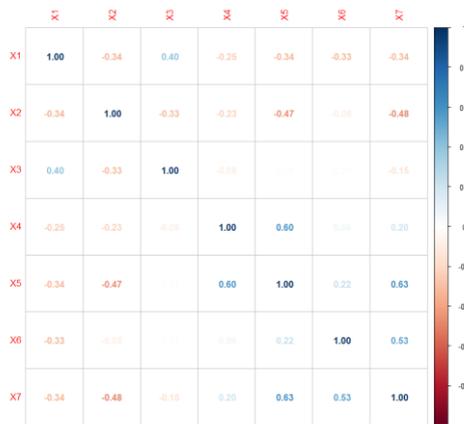
1. Atribut-atribut yang digunakan sebagai dasar pengelompokan provinsi-provinsi di Indonesia distandardisasi terlebih dahulu.

2. Memeriksa korelasi antar atribut.
3. Menentukan jumlah kluster k menggunakan metode *average silhouette* baik menggunakan algoritma *k-means* dan *k-medoids*.
4. Mengalokasikan setiap data atau objek ke kluster terdekat yang telah ditentukan menggunakan persamaan ukuran jarak. Dalam penelitian ini, baik algoritma *k-means* maupun *k-medoids* akan menggunakan jarak Euclidean agar dapat dibandingkan hasil analisis kluster.
5. Proses alokasi setiap data/objek ke kluster terdekat merupakan proses iteratif hingga pusat kluster yang berupa *centroid* dan *medoid* tidak berubah lagi.
6. Melakukan pemilihan algoritma terbaik menggunakan validasi internal yang mencakup *connectivity*, *dunn index*, dan *silhouette*.
7. Melakukan interpretasi pada masing-masing kluster (*cluster profiling*) menggunakan algoritma terbaik yang terpilih.

3. Hasil dan Diskusi

3.1 Pemeriksaan Korelasi Antar Atribut

Multikolinieritas menjadi masalah serius ketika korelasi antara dua variabel independen melebihi 0,8 [16]. Khusus dalam analisis kluster, korelasi dapat terjadi pada atribut-atribut yang membentuk sebuah kluster. Gambar 2 menunjukkan bahwa tidak ada korelasi antara dua atribut yang melebihi 0,8. Sehingga, atribut-atribut yang dipilih sebelumnya sudah dapat digunakan dalam proses pembentukan kluster menggunakan algoritma *k-means* dan *k-medoids*.



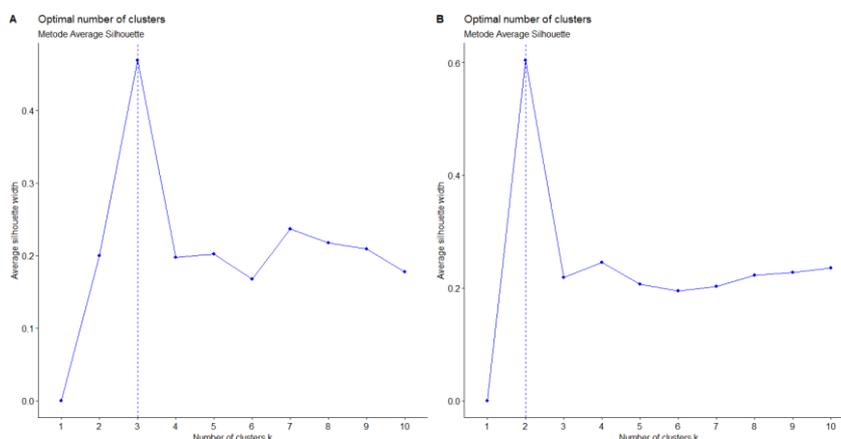
Sumber: Hasil Pengolahan

Gambar 1. Korelasi antar atribut

3.2 Penentuan Jumlah Kluster Optimum

Penentuan jumlah kluster optimum menggunakan metode *average silhouette* dan diterapkan pada algoritma *k-means* dan *k-medoids*. Berdasarkan gambar 2, nilai rata-rata

koefisien *silhouette* mencapai titik maksimum pada jumlah kluster sebanyak tiga dengan menggunakan algoritma *k-means* dan sebanyak dua dengan menggunakan algoritma *k-medoids*. Titik maksimum tersebut menandakan jumlah kluster optimal yang dapat dibentuk dengan masing-masing algoritma.



Sumber: Hasil Pengolahan

Gambar 2. Jumlah kluster optimum A) *k-means* B) *k-medoids*

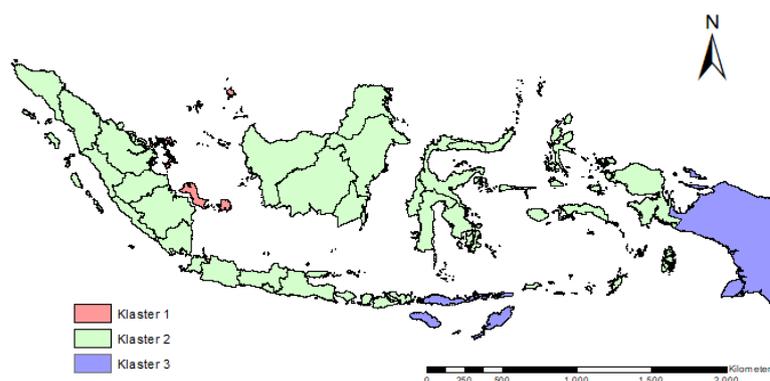
3.3 Hasil Analisis Kluster *K-Means* dan *K-Medoids*

Hasil analisis kluster dengan algoritma *k-means* menghasilkan tiga kluster provinsi di Indonesia berdasarkan indikator perumahan rumah tangga. Terlihat pada tabel 2 dan gambar 3, kluster satu mengelompokkan 3 provinsi, kluster dua mengelompokkan 29 provinsi, dan kluster tiga mengelompokkan 2 provinsi.

Tabel 2. Hasil analisis kluster menggunakan algoritma *k-means*

Kluster	Provinsi
1	Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta
2	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
3	Nusa Tenggara Timur, Papua

Sumber: Hasil Pengolahan



Sumber: Hasil Pengolahan

Gambar 3. Visualisasi analisis kluster menggunakan algoritma *k-means* berdasarkan indikator perumahan rumah tangga 2020

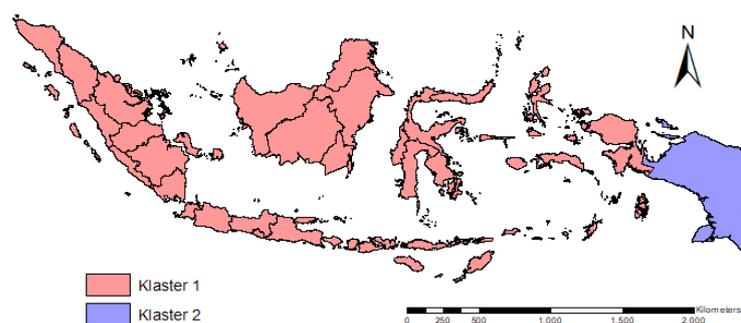
Sementara itu, sebagai bentuk pembandingan, hasil analisis kluster menggunakan algoritma *k-medoids* menghasilkan dua kluster provinsi di Indonesia berdasarkan indikator perumahan rumah tangga namun muatan masing-masing kluster yang dihasilkan cukup berbeda dibandingkan *k-means* sebelumnya. Terlihat pada tabel 3 dan gambar 4, kluster satu mengelompokkan 33 provinsi dan kluster dua hanya mengelompokkan 1 provinsi.

Tabel 3. Hasil analisis kluster menggunakan algoritma *k-medoids*

Kluster	Provinsi
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat
2	Papua

Sumber: Hasil Pengolahan, 2022

Klasterisasi provinsi-provinsi di Indonesia menggunakan algoritma *k-means* dan *k-medoids* cukup memberikan hasil yang cukup berbeda. Hal ini diakibatkan oleh proses iterasi dalam pemilihan pusat kluster kedua metode tersebut. Terlihat bahwa Provinsi Papua baik menggunakan algoritma *k-means* maupun *k-medoids* masuk pada kluster tiga. Hal ini mengindikasikan bahwa Provinsi Papua merupakan provinsi yang layak diperhatikan berdasarkan indikator perumahan rumah tangganya.



Sumber: Hasil Pengolahan

Gambar 4. Visualisasi analisis kluster menggunakan algoritma *k-medoids* berdasarkan indikator perumahan rumah tangga 2020

3.4 Penentuan Algoritma Kluster Terbaik

Sebelum menginterpretasikan setiap kluster, algoritma yang dapat mengelompokkan provinsi dengan baik harus terlebih dahulu ditentukan. Penentuan algoritma kluster terbaik menggunakan ukuran validitas internal yaitu *connectivity*, *dunn index*, dan koefisien *silhouette*. Tabel 4 menunjukkan bahwa algoritma *k-means* memiliki nilai *connectivity* dan *dunn index* yang lebih besar dibandingkan algoritma *k-medoids*. Sedangkan jika dilihat dari nilai *silhouette*, algoritma *k-means* memiliki nilai yang lebih rendah dibandingkan algoritma *k-medoids*. Hal ini mengindikasikan bahwa tidak ada algoritma klustering yang bisa digunakan secara universal dalam pengelompokan provinsi-provinsi di Indonesia berdasarkan indikator perumahan rumah tangga. Sehingga baik, *k-means* maupun *k-medoids* tetap dapat digunakan untuk pengelompokan provinsi-provinsi di Indonesia berdasarkan indikator perumahan rumah tangga.

Tabel 4. Perbandingan algoritma *k-means* dan *k-medoids*

Algoritma	<i>Connectivity</i>	<i>Dunn Index</i>	<i>Silhouette</i>
K-means	5,8579	0,7244	0,5925
K-medoids	2,9290	0,6887	0,6045

Sumber: Hasil Pengolahan

3.5 Interpretasi Kluster Menggunakan Algoritma Terbaik

Interpretasi hasil masing-masing kluster dilakukan dengan menggunakan algoritma *k-means*. Selain dilatarbelakangi kriteria pemilihan algoritma terbaik sebelumnya, Algoritma *k-means* dipilih sebab dilatarbelakangi hasil analisis kluster yang lebih variatif dibandingkan algoritma *k-medoids*. Berdasarkan tabel 5, kluster satu mencirikan provinsi-provinsi yang karakteristiknya indikator rumah tangganya masih perlu diperhatikan, khususnya pada status kepemilikan hunian miliki sendiri, luas hunian per

kapita di bawah 7,2 m², dan jenis atap terluas. Hal ini karena nilai-nilai pada indikator tersebut masih di atas atau di bawah dari angka nasional. Kemudian klaster dua mencirikan provinsi-provinsi yang karakteristik indikator rumah tangganya patut untuk dipertahankan. Hal ini karena hanya ada satu nilai indikator yaitu luas hunian per kapita di bawah 7,2 m² yang di atas angka nasional. Terakhir pada klaster tiga, klaster tersebut mencirikan provinsi-provinsi yang karakteristik indikator perumahan rumah tangganya sangat butuh perhatian. Hal ini karena hampir keseluruhan indikator perumahan pada klaster provinsi tersebut masih di atas dan di bawah dari angka nasional. Klaster tiga menunjukkan bahwa rata-rata rumah tangga di provinsi tersebut cenderung memiliki hunian sendiri, namun hunian tersebut masih belum dapat dikategorikan sebagai hunian layak, aman, dan terjangkau. Berbeda halnya dengan klaster satu, klaster satu memiliki hunian yang cukup layak, namun kepemilikan hunian milik sendiri umumnya masih cukup rendah di klaster provinsi tersebut. Hasil analisis klaster juga menunjukkan bahwa secara umum provinsi-provinsi di Indonesia yang dikelompokkan berdasarkan indikator perumahan rumah tangganya berada pada klaster dua. Artinya bahwa rumah tangga di Indonesia secara rata-rata telah memiliki hunian yang layak, aman, dan terjangkau.

Tabel 5. Hasil pencirian masing-masing klaster menggunakan algoritma *k-means*

Atribut	Klaster 1	Klaster 2	Klaster 3	Angka Acuan (Nasional)
Status hunian milik sendiri	65,58*	80,03	85,22	80,10
Luas hunian per kapita $\leq 7,2$ m ²	11,8068*	8,4597*	22,7599*	7,85
Atap terluas bukan ijuk/lainnya	49,1133*	94,2717	83,1050*	89,10
Dinding terluas bukan bambu/lainnya	99,6819	98,5638	82,2847*	98,20
Lantai terluas bukan tanah	99,8	97,780	78,9650*	95,88
Sumber air minum layak	88,4356	85,9261	73,3023*	90,21
Akses terhadap sanitasi layak	91,6033	80,3065	55,005*	79,53

Sumber: Hasil Pengolahan, 2022

* : indikator perumahan yang menjadi perhatian jika dibandingkan dengan angka nasional

Klaster satu yang mencakup Provinsi Kepulauan Bangka Belitung, Kepulauan Riau, dan DKI Jakarta memiliki masalah pada salah satu atau keseluruhan ketiga indikator perumahan rumah tangga yaitu kepemilikan hunian milik sendiri, luas hunian per kapita, dan atap terluas. Menurut BPS [8], Provinsi Kepulauan Bangka Belitung memiliki persentase rumah tangga menurut status kepemilikan hunian milik sendiri sebesar 85,33 persen, sedangkan sisanya sebesar 14,67 persen bukan milik sendiri yang mencakup kontrak/sewa, bebas sewa, dan dinas. Kemudian Provinsi Kepulauan Riau memiliki persentase rumah tangga menurut status kepemilikan hunian milik sendiri sebesar 66,37

persen, sedangkan sisanya sebesar 33,63 persen bukan milik sendiri yang mencakup kontrak/sewa, bebas sewa, dan dinas. Terakhir Provinsi DKI Jakarta memiliki persentase rumah tangga menurut status kepemilikan hunian milik sendiri sebesar 45,04 persen, sedangkan sisanya sebesar 54,96 persen bukan milik sendiri yang mencakup kontrak/sewa, bebas sewa, dan dinas. Khusus untuk DKI Jakarta yang persentase status kepemilikan rumah tangga milik sendiri tidak mendominasi kepemilikan rumah tangga disebabkan karena wilayah DKI Jakarta merupakan wilayah yang padat penduduk sehingga membuat tingginya permintaan akan hunian tempat tinggal dan harga hunian tersebut yang mahal. Sementara kedua provinsi lainnya memiliki persentase status kepemilikan hunian rumah tangga milik sendiri yang masih mendominasi, walaupun pada Provinsi Kepulauan Riau nilainya masih cukup rendah. Kemudian kalau dilihat dari indikator jenis atap terluas yang menjadi masalah cukup besar di klaster ini, terjadi karena hunian yang ditempati rumah tangga baik di Provinsi Kepulauan Bangka Belitung, Kepulauan Riau, maupun DKI Jakarta memiliki atap hunian yang terluas yaitu asbes dengan nilai cukup besar yaitu 54,25 persen untuk wilayah Provinsi Kepulauan Bangka Belitung, 38,72 persen untuk wilayah Provinsi Kepulauan Riau dan 58,09 persen untuk wilayah Provinsi DKI Jakarta [8]. BPS sejak tahun 2019 tidak memasukkan kategori asbes sebagai atap terluas sebagai salah satu kriteria hunian yang layak, aman, dan terjangkau. Sehingga berdampak pada penilaian kriteria hunian yang layak pada kedua provinsi di dalam klaster ini.

Klaster tiga yang mencakup Provinsi Nusa Tenggara Timur dan Papua sebenarnya memiliki masalah utama pada indikator perumahan rumah tangga yaitu luas hunian per kapita di bawah 7,2 m², jenis lantai terluas, jenis sumber air minum, dan akses terhadap sanitasi layak. Sedangkan indikator jenis atap terluas dan jenis dinding terluas nilainya sudah cukup baik, tetapi tetap perlu adanya peningkatan pada kedua indikator tersebut agar dapat melebihi angka nasional. BPS [8] menyebutkan bahwa pada Provinsi Nusa Tenggara Timur dan Papua masih memiliki angka yang cukup besar pada jenis lantai terluas tanah hunian rumah tangga yaitu masing-masing sebesar 18,16 persen dan 20,65 persen. Kemudian pada indikator sumber air minum, Provinsi Papua sebagai penyumbang terbesar jenis sumber air minum yang tidak layak pada klaster tiga yaitu sebesar 37,27 persen. Terakhir, indikator yang cukup rendah capaiannya yaitu akses terhadap sanitasi layak menjadi masalah bagi kedua provinsi yang ada di dalam klaster tiga ini. BPS [8] juga menjelaskan bahwa hunian rumah tangga di Provinsi Nusa Tenggara Timur yang memiliki akses terhadap sanitasi yang tidak layak sebesar 30,3 persen sementara Provinsi Papua yang memiliki akses terhadap sanitasi yang tidak layak sebesar 59,69 persen. Kedua provinsi tersebut merupakan provinsi dengan rangking yang paling atas pada indikator akses terhadap sanitasi yang tidak layak pada tahun 2020.

4. Kesimpulan

Berdasarkan hasil analisis kluster yang dilakukan menurut indikator perumahan rumah tangga, terdapat tiga kluster provinsi-provinsi di Indonesia menggunakan algoritma *k-means* dan terdapat dua kluster provinsi-provinsi di Indonesia menggunakan algoritma *k-medoids*. Namun berdasarkan indikator pemilihan algoritma kluster terbaik, *k-means* merupakan algoritma yang lebih baik untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator perumahan rumah tangga dibandingkan *k-medoids*. Algoritma *k-means* dipilih selain disebabkan oleh indikator pemilihan algoritma terbaik, tetapi juga disebabkan hasil analisis kluster yang dihasilkan lebih variatif. Kemudian berdasarkan hasil pencirian kluster menggunakan algoritma *k-means*, kluster satu mencirikan provinsi-provinsi yang karakteristiknya indikator rumah tangganya masih perlu diperhatikan, khususnya pada status kepemilikan hunian miliki sendiri, luas hunian per kapita di bawah 7,2 m², dan jenis atap terluas. Kluster dua mencirikan provinsi-provinsi yang karakteristik indikator rumah tangganya patut untuk dipertahankan sebab hanya ada satu nilai indikator yaitu luas hunian per kapita di bawah 7,2 m² yang di atas angka nasional. Sementara itu kluster tiga mencirikan provinsi-provinsi yang karakteristik indikator perumahan rumah tangganya sangat butuh perhatian sebab hampir keseluruhan indikator perumahan pada kluster provinsi tersebut masih di atas dan di bawah dari angka nasional. Penelitian ini menjadi suatu indikasi awal bagi pemerintah dalam meningkatkan persentase hunian layak, aman, dan terjangkau bagi masyarakat. Pemerintah diharapkan dapat menangani masalah-masalah indikator perumahan rumah tangga di Indonesia dengan mempelajari secara lebih dalam mengenai apa-apa saja indikator dan wilayah yang perlu ditangani terutama dengan melihat pengelompokan wilayah provinsi yang telah dilakukan pada penelitian ini.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih banyak kepada Bapak Dr. Rindang Bangun Prasetyo, S.S.T., M.Si., yang telah memberikan saran sehingga tulisan ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] PUPR, "Program Sejuta Rumah Per Juli Capai 515.107 Unit," 16 Agustus 2021. [Online]. Available: <https://perumahan.pu.go.id/news/program-sejuta-rumah-per-juli-capai-515107-unit>.
- [2] Kompas.com, "Pemerintah Targetkan 11 Juta Rumah Tangga Tempati Rumah Layak Huni hingga 2024," 1 Juli 2021. [Online]. Available: <https://www.kompas.com/properti/read/2021/07/01/060000721/pemerintah-targetkan-11-juta-rumah-tangga-tempati-rumah-layak-huni>.
- [3] BPS, Statistik Perumahan dan Permukiman 2019, Jakarta: BPS, 2020.

- [4] PPDPP, “Di Tengah Pandemi COVID-19, Realisasi Program Sejuta Rumah TA 2020 Capai 965.217 Unit,” 8 Januari 2021. [Online]. Available: <https://ppdpp.id/di-tengah-pandemi-covid-19-realisasi-program-sejuta-rumah-ta-2020-capai-965-217-unit/>.
- [5] BPS, “Persentase Rumah Tangga yang Memiliki Akses Terhadap Hunian Yang Layak Dan Terjangkau Menurut Daerah Tempat Tinggal (Persen), 2018-2020,” 26 Januari 2021. [Online]. Available: <https://www.bps.go.id/indicator/29/1242/1/persentase-rumah-tangga-yang-memiliki-akses-terhadap-hunian-yang-layak-dan-terjangkau-menurut-daerah-tempat-tinggal.html>.
- [6] W. Wijayanti, I. Rahmi dan F. Yanuar, “Penggunaan Metode Fuzzy C-Means Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesehatan Lingkungan,” *Jurnal Matematika UNAND*, vol. 10, no. 1, pp. 129-136, 2021.
- [7] V. A. Lathuhimallo, M. W. Talakua dan Z. A. Leleury, “Analisis Clustering Untuk Pengelompokan Kabupaten/Kota Berdasarkan Indikator Kesejahteraan Rakyat di Wilayah Provinsi Maluku,” *Jurnal Matematika, Statistika dan Terapannya*, vol. 2, no. 1, pp. 127-138, 2021.
- [8] BPS, *Indikator Perumahan dan Kesehatan Lingkungan 2020*, Jakarta: BPS, 2020.
- [9] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*, Gresik: Penerbit ANDI, 2013.
- [10] D. Jollyta, W. Ramdhan dan M. Zarlis, *Konsep Data Mining dan Penerapan*, Pekanbaru: deepublish, 2020.
- [11] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data Edisi Revisi*, Bandung: INFORMATIKA, 2019.
- [12] B. Wira, A. E. Budianto dan A. S. Wiguna, “Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang,” *Jurnal Terapan Sains & Teknologi*, vol. 1, no. 3, pp. 54-69, 2019.
- [13] J. Han, M. Kamber dan J. Pei, *Data Mining Concepts and Techniques Third Edition*, Waltham: Elsevier Inc., 2012.
- [14] A. Kassambara, *Practical Guide To Cluster Analysis in R*, sthda.com, 2017.
- [15] A. C. Benabdellah, A. Benghabrit dan I. Bouhaddou, “A Survey of Clustering Algorithms For an Industrial Context,” *Procedia Computer Science*, vol. 148, no. 2019, pp. 291-302, 2019.
- [16] D. N. Gujarati dan D. C. Porter, *Basic Econometrics Fifth Edition*, New York: The McGraw-Hill Companies, 2009.